




Deep-learning-based whole-brain imaging at single-neuron resolution

KEFU NING,^{1,3} XIAOYU ZHANG,^{1,3} XUEFEI GAO,^{1,2} TAO JIANG,² HE WANG,¹ SIQI CHEN,¹ ANAN LI,^{1,2} AND JING YUAN^{1,2,*} 

¹Britton Chance Center for Biomedical Photonics, Wuhan National Laboratory for Optoelectronics, MoE Key Laboratory for Biomedical Photonics, School of Engineering Sciences, Innovation Institute, Huazhong University of Science and Technology, Wuhan 430074, China

²HUST-Suzhou Institute for Brainsmatics, JITRI Institute for Brainsmatics, Suzhou 215000, China

³These authors contributed equally to this work

*yuanj@hust.edu.cn

Abstract: Obtaining fine structures of neurons is necessary for understanding brain function. Simple and effective methods for large-scale 3D imaging at optical resolution are still lacking. Here, we proposed a deep-learning-based fluorescence micro-optical sectioning tomography (DL-fMOST) method for high-throughput, high-resolution whole-brain imaging. We utilized a wide-field microscope for imaging, a U-net convolutional neural network for real-time optical sectioning, and histological sectioning for exceeding the imaging depth limit. A 3D dataset of a mouse brain with a voxel size of $0.32 \times 0.32 \times 2 \mu\text{m}$ was acquired in 1.5 days. We demonstrated the robustness of DL-fMOST for mouse brains with labeling of different types of neurons.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Decades of studies have shown that neural connectivity holds the key to brain function [1,2]. Deciphering neural circuits across the entire brain is central to better understanding of the brain mechanism. With the advancement of fluorescent labeling techniques, we can focus on specific neural circuits [3,4]. However, neurons have multi-scale characteristics. The axon from the cell body is about $1 \mu\text{m}$ in diameter, but its length can extend to a few millimeters [5,6]. Three-dimensional imaging with sub-micron resolution at such a large span is impossible for conventional optical microscopy. To address this challenge, several whole-brain optical imaging methods have been developed.

Optical clearing methods allow light to penetrate deep into the tissue with less absorption and scattering, thereby extending the imaging depth [7]. Combined with light-sheet microscopy, a full-volume mouse brain dataset can be obtained within hours [8–11]. However, the resolutions in these cases are limited due to the use of low-magnification objective lenses with long working distance. Since the brain cannot be completely transparent, a scattering problem exists when imaging deep into the brain, making it impossible to achieve uniform resolution throughout the whole volume.

Another way to overcome the depth limitation of optical imaging is to combine optical microscopy with automatic histological sectioning [12]. The data acquisition of Golgi-stained mouse brain at optical resolution has been demonstrated using micro-optical sectioning tomography (MOST) technologies employing imaging and sectioning simultaneously [13]. A fluorescence MOST (fMOST) system that substituted confocal fluorescence imaging for reflection imaging showed long-range projections of single axons in the Thy1 transgenic (Thy1-eYFP-H and Thy1-GFP-M) mouse brains [14]. Direct imaging of the cut tissues on the knife edge requires very strict cutting quality. The surface roughness of cutting directly affects the quality of data and any defect can lead to knife breakage and data loss. To avoid this problem, block-face imaging before tissue sectioning has been adopted to separate the imaging and sectioning parts, and

different optical-sectioning mechanisms have been introduced in the imaging part to eliminate strong background fluorescence from deep tissue. Ragan et al. adopted two-photon imaging to avoid imaging on the top surface and applied a vibratome for interval sectioning to develop serial two-photon tomography (STP) and acquire a whole-brain dataset at 50 μm axial interval within 1 day [15]. Zheng et al. employed an acoustic-optic deflector for stably and durably inertialess scanning to achieve whole-brain imaging with axonal resolution in 8–10 days [16]. The Mouselight platform improved the STP system by using a high-speed resonant scanner and integrating it with tissue clearing to provide another solution for tracing long-range projections with a data acquisition time of 7 days [17,18]. Gong et al. introduced structured illumination microscopy (SIM) [19] to improve imaging throughput and achieved a brain-wide acquisition at a voxel size of $0.32 \times 0.32 \times 2 \mu\text{m}$ in 77 h [20]. In this case, a digital micro-mirror device (DMD) had to refresh more than one million times for imaging of a mouse brain. Recently, block-face serial microscopy tomography (FAST) employed spinning disk scanning to further shorten the imaging time of a mouse brain to 2.4 hours with a sacrificed resolution of $0.7 \times 0.7 \times 5 \mu\text{m}$ [21]. By harnessing oblique light-sheet imaging, the high-throughput light-sheet tomography platform (HLTP) was able to image a whole mouse brain in 5 h at a voxel size of $1.30 \times 1.30 \times 0.92 \mu\text{m}$ [22]. However, all these optical-sectioning methods not only improve the data quality, but also increase the complexity of the system configuration and the difficulty of stable operation. Thus, it is highly desirable to design a simplified instrument that maintains high performance.

Deep learning technology [23] has been widely used in the field of biomedical imaging, especially for image reconstruction [24]. A data-driven approach based on deep learning has the potential to break the constraints of hardware conditions and integrate the advantages of different optical systems [25]. Here, we propose a deep-learning-based fluorescence MOST (DL-fMOST) method to reveal 3D distribution of specific labels by the combination of wide-field (WF) imaging, histological sectioning, and deep learning prediction. To our best knowledge, this is the first time that a deep learning method is used to support whole-brain imaging. Our group has reported the demonstration of deep learning for optical sectioning in traditional optical microscopy [26]. However, the network was very simple which was prone to underfitting when applied to a whole-brain dataset containing millions of images with diverse image features. Additionally, the inference speed was too low to realize on-line processing. Even for post-processing, it would take nearly a month to predict the mouse whole-brain dataset. In this study, we employed a wider and deeper network architecture to better learn complex image features throughout the whole-brain dataset. We also redesigned the reconstruction process to achieve real-time image reconstruction. To validate the reliability of this method, we acquired mouse brain datasets of different types of neurons and projections. The performance of DL-fMOST was quantitatively tested using both the structural similarity index (SSIM) [27] and root mean squared error (RMSE). We also performed cell counting and neuronal morphology reconstruction to further quantify the image quality of DL-fMOST. Our results demonstrated that DL-fMOST could potentially facilitate neuroscience research, especially in revealing and visualizing the brain-wide distributions and projection patterns of type-specific neurons.

2. Materials and methods

2.1. DL-fMOST

A brief system diagram is shown in Fig. 1(a). The imaging part was a typical wide-field fluorescence microscope without complex hardware set up for optical sectioning. This imaging part took advantages of wide-field imaging for high throughput and effective cost. An excitation beam from a mercury lamp source (X-Cite exacte, Lumen Dynamics, Canada) was transmitted through a tube lens (TL1, $f = 150 \text{ mm}$) and filtered by an excitation filter (EX, FF02-482, Semrock Inc., USA). Then, the excitation beam was reflected on a dichroic mirror (DM, FF495-Di03, Semrock Inc., USA) and focused on a sample surface through an objective lens (XLUMPLFLN

20XW, Olympus, Japan). A piezoelectric translational stage (PZT, P-725 PIFOC Long-Travel Objective Scanner, PI GmbH, Germany) moved the objective for axial scanning. The excited fluorescence was filtered by an emission filter (EM, FF02-520, Semrock Inc., USA) and detected by a scientific complementary metal oxide semiconductor (sCMOS) camera (ORCA-Flash 4.0, Hamamatsu Photonics K.K., Japan). The coronal images of the sample were obtained by mosaic scanning. After imaging, a 3D translation stage (ABL20020-ANT130-AVL125, Aerotech, Inc., USA) moved the sample to the microtome (Diatome AG, Switzerland) for removal of the imaged tissue. The process of imaging and sectioning was repeated until the entire volume was acquired. To remove the defocus background fluorescence, we used a well-trained convolutional neural network (CNN) to process the acquired WF image to the corresponding optical-sectioning images in real time (Fig. 1(b)). All the animal experiments followed procedures approved by the Institutional Animal Ethics Committee of Huazhong University of Science and Technology.

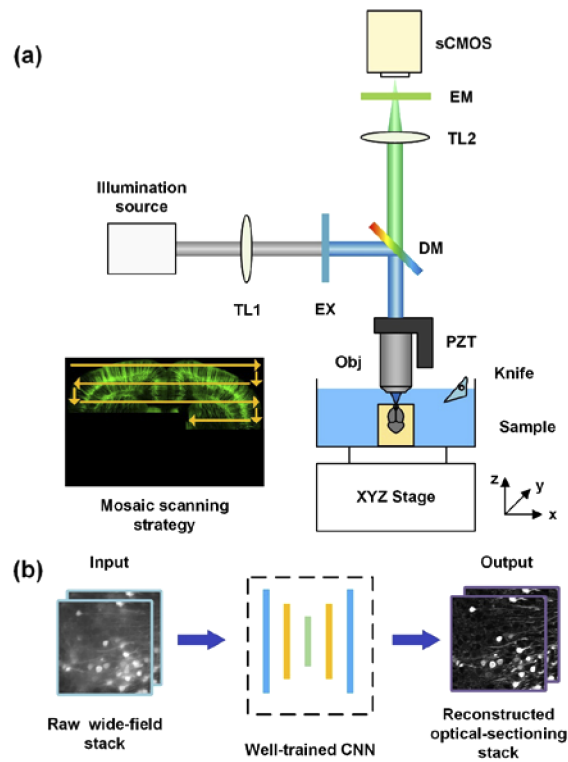


Fig. 1. Principle of DL-fMOST. (a) System configuration and imaging strategy. TL, tube lens; EX, excitation filter; DM, dichroic mirror; PZT, piezoelectric translational stage; Obj, objective; EM, emission filter. (b) Optical sectioning enabled by the trained convolutional neural network.

2.2. Data preparation

In previous studies [28,29], researchers usually spent a lot of time and effort for accurate registration. Here, we used the WVT system [20] to acquire raw data to simultaneously reconstruct co-located WF images and SIM images, without extra data registration. The WVT system utilized a DMD for fast structured light illumination modulation and acquired three equal phase-stepped raw images in each FOV. Then the WF image and the corresponding SIM

optical-sectioned image can be reconstructed as follows [19]:

$$\begin{cases} I_{WF} = \frac{1}{3}(I_1 + I_2 + I_3) \\ I_{SIM} = [(I_1 - I_2)^2 + (I_2 - I_3)^2 + (I_3 - I_1)^2]^{1/2} \end{cases} \quad (1)$$

where I_1 , I_2 , and I_3 denote three raw images with the phases of 0, $2/3\pi$, and $4/3\pi$, respectively. The SIM reconstruction provided a ground truth for optical sectioning. We imaged a resin-embedded 8-week-old Thy1-GFP M-line [30] mouse brain with a voxel size of $0.32 \times 0.32 \times 2 \mu\text{m}$ and reconstructed its WF and SIM dataset. All procedures related to sample preparation have been previously described [20]. Neuron morphology, cell density, and signal-to-noise-ratio (SNR) in different brain regions are quite different. Figure 2 shows five pairs of randomly-selected typical images with various image features. We randomly extracted three images at $100 \mu\text{m}$ intervals and obtained a total of 300 images to build a complete and unbiased training set, using 200 pairs for training and the others for validation. After random cropping, 9800 image patches were used for training and 4900 patches for validating. The crop size was 256×256 pixels. We found more training data worked similarly well but increased training time.

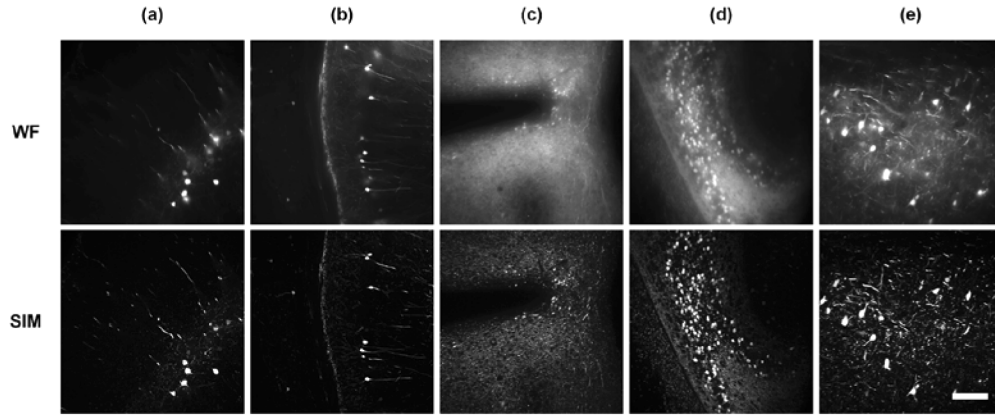


Fig. 2. Typical training data for DL-fMOST. The WF images and the corresponding SIM images are taken from the same dataset at the positions of Bregma 0.86 mm (a), -0.22 mm (b), -1.22 mm (c), -2.54 mm (d), and -3.80 mm (e). Scale bar: $100 \mu\text{m}$.

Five image sets were prepared for testing, which included a whole-brain dataset from a Thy1-GFP-M transgenic mouse (test set 1) and four virus-labeled whole-brain datasets with various axon projections (test set 2-5). For virus injection, the mouse brain was first injected with adeno-associated virus (AAV) helper mixtures (mixed with rAAV2/9-Ef1 α -DIO-BFP-2a-TVA-WPRE-pA and rAAV2/9-Ef1 α -DIO-RG-WPRE-pA as the ratio of 1:2) to provide the receptor for EnvA-coated rabies virus (RV). After 3 weeks, the RV (RV- Δ G-EnVA-EGFP) was injected at the same site. Specifically, the virus labeling included a Thy1-Cre mouse injected in the primary somatosensory cortex, barrel field (150 nl AAV and 300 nl RV, test set 2), a vGluT2-Cre mouse injected in the substantia nigra, compact part (100 nl AAV and 200 nl RV, test set 3), a Fezf2-2A-CreER mouse injected in the secondary auditory cortex, dorsal area (100 nl AAV and 300 nl RV, test set 4), and a Fezf2-2A-CreER mouse injected in the secondary auditory cortex, ventral area (100 nl AAV and 300 nl RV, test set 5). All the viruses were produced by BrainVTA (Wuhan, China). We applied three WVT systems to obtain these raw data and the illumination power and exposure time were different according to signal intensity. Specifically, the training set and test set 1 were obtained by system 1, test set 2 and 4 were acquired by system 2 and test set 3 and 5 were obtained by system 3. All these systems have the same configuration.

2.3. Network structure and training

The CNN architecture for fast optical sectioning is illustrating in Fig. 3. In our previous work [26], a relatively simple network was employed to allow network training by only one pair of images. However, such a small network was not able to represent the complex features in the whole brain dataset. Additionally, as the output size of the network was 1×1 , it was necessary to go through every pixel to reconstruct a single image, which was very time-consuming. Here, we built a wider and deeper network with structure similar to U-net, which has been widely used in super-resolution imaging [31,32], label-free prediction [33,34], and image denoising [35,36] for its efficiency and practicality. The network consists of 16 convolutional layers (zero-padded), and both the input and output sizes are 256×256 pixels. It contains an encoding path and a decoding path, which are symmetrical to each other. In the encoding path, 8 convolution layers of 4×4 kernel and 2×2 stride gradually down-sample the input image patch from 256×256 to 1×1 pixels. The decoding path uses 8 transposed convolution layers (4×4 kernel and 2×2 stride) to recover the image from 1×1 to 256×256 pixels. The activation layers in the encoder are leaky rectified linear units (ReLU) [37] with slope 0.2, while those in the decoder are ReLU [38]. There is a skip connection between each layer i and layer $n - i$, where n indicates the total number of convolutional layers. The skip connection simply concatenates all features between each layer i and layer $n - i$. This operation allows low-level information to pass directly to high-level layers, which is useful in end-to-end image transformation. In order to make the training process more stable and have better generalization performance, a BatchNormalization layer [39] and a Dropout layer [40] are employed in the network.

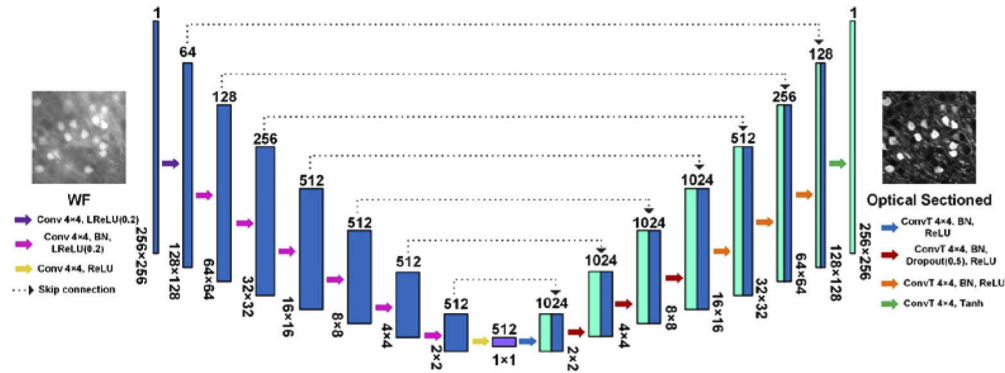


Fig. 3. Network architecture. The numbers on the top of the block indicate the output channels of each convolutional layer, and the numbers on the bottom of the block represent the size of the feature map. The operations are represented by arrows of different colors.

The network training process iteratively optimizes the parameters of the network through a back-propagation algorithm. In order to measure the distance between the network output and the ground truth, we used mean absolute error (MAE) as the loss function of the network:

$$L_{MAE} = \frac{1}{KHW} \sum_{k=1}^K \sum_{i=1}^H \sum_{j=1}^W |\psi_{\theta}^k(i,j) - P(i,j)| \quad (2)$$

where Ψ_{θ} represents the network reconstructed image, P is the ground truth image, and K , H , W denote the mini-batch size, image height, and image width, respectively. Compared to mean squared error, MAE enables better visual effects in the restored image [41].

The network parameters were optimized by the Adam optimizer [42] with learning rate 0.0002. It was trained for 200 epochs, using data augmentation such as random flipping to enhance its

robustness, and the learning rate was halved after every 50 epochs. The dropout rate was set to 0.5 and the mini-batch size was 16. The network was trained and tested on a workstation (Precision T7920 Tower, Dell Inc., USA) with dual-core 2.1 GHz and 256 GB of RAM using two Nvidia Titan XP GPUs. It was built using an open-source deep learning package, PyTorch (Version 1.0.1) [43] on Python 3.6.5. It took approximately 15 h for training.

2.4. Real-time processing

An optimized inference process was proposed to enable real-time optical sectioning reconstruction by the trained CNN. Specifically, the steps were as follows: (a) Stitch the acquired WF images into a complete coronal image. (b) Crop the coronal image into sub-images of 256×256 pixels and store the images in a queue. (c) Assign two separate threads to the two GPUs, and let them predict the images waiting in the queue with the largest batch size, respectively. (d) Stitch the sub-images predicted by the network into an entire coronal image. After optimization, the inference time of the largest coronal plane in the hippocampus (about 32684×20380 pixels) was 14.6 s. Sectioning time for a coronal plane was about 24 s. This means that our processing speed allows reconstruction of the optical-sectioning image for each acquired image in real time during the time window of tissue sectioning. A typical whole-brain dataset acquired by the WVT system contained 560,000 mosaic images; real-time processing effectively speeded up the process for whole-brain data reconstruction and analysis.

2.5. Performance criteria

The performance of our network was evaluated by the following three criteria. RMSE is the simplest and most widely used quality metric [44]. A lower RMSE value indicates the image with less distortion. It is defined as

$$RMSE = \left[\frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W (M(i,j) - N(i,j))^2 \right]^{1/2} \quad (3)$$

where M denotes the network output, N denotes the corresponding ground truth, and H and W are the height and width of the image, respectively. The pixel values were normalized to 0 to 1 before calculation.

SSIM is another widely accepted quality assessment that combines luminance comparison, contrast comparison, and structure comparison to give the final similarity measure [27]. The output value of SSIM is between 0 and 1. A larger SSIM index indicates higher fidelity. Equation (4) shows how SSIM is calculated:

$$SSIM = \frac{(2\mu_M\mu_N + C_1)(2\sigma_{MN} + C_2)}{(\mu_M^2 + \mu_N^2 + C_1)(\sigma_M^2 + \sigma_N^2 + C_2)} \quad (4)$$

where M and N denote the network output and corresponding ground truth; μ_M and μ_N are the mean value of M and N , respectively; σ_M and σ_N are the standard deviations of M and N , respectively; and σ_{MN} is the covariance of M and N . C_1 and C_2 are constants to avoid the denominator close to zero.

We quantified the SNRs of SIM image and CNN image by Eq. (5) [32]:

$$SNR = \left| \frac{p - b}{\sigma_b} \right| \quad (5)$$

where p is peak signal of the FOV and b and σ_b are the mean value and the standard deviation of the background, respectively. The signal-to-background ratio (SBR) was employed to quantify the optical-sectioning abilities of different methods, which was defined by Eq. (6) [26]:

$$SBR = \frac{p}{b}. \quad (6)$$

3. Results

3.1. Whole-brain imaging

To validate DL-fMOST for whole-brain imaging, we reconstructed the entire dataset of test set 1 by feeding the WF raw images into the network trained with Thy1-GFP-M whole brain data in Section 2.2. Figure 4 shows the maximum intensity projections (MIP) of three coronal sections at equal intervals of 2 mm by SIM reconstruction and CNN prediction. There were significant differences in neuron distribution and morphology in the whole brain. In Fig. 4(a), the neural fibers were densely and evenly distributed, whereas in Fig. 4(b), the neuron distribution was relatively sparse. In Fig. 4(c), we can observe remarkably different cell density in different brain regions. The neurons were tightly spaced in the retrohippocampal region (marked by orange)

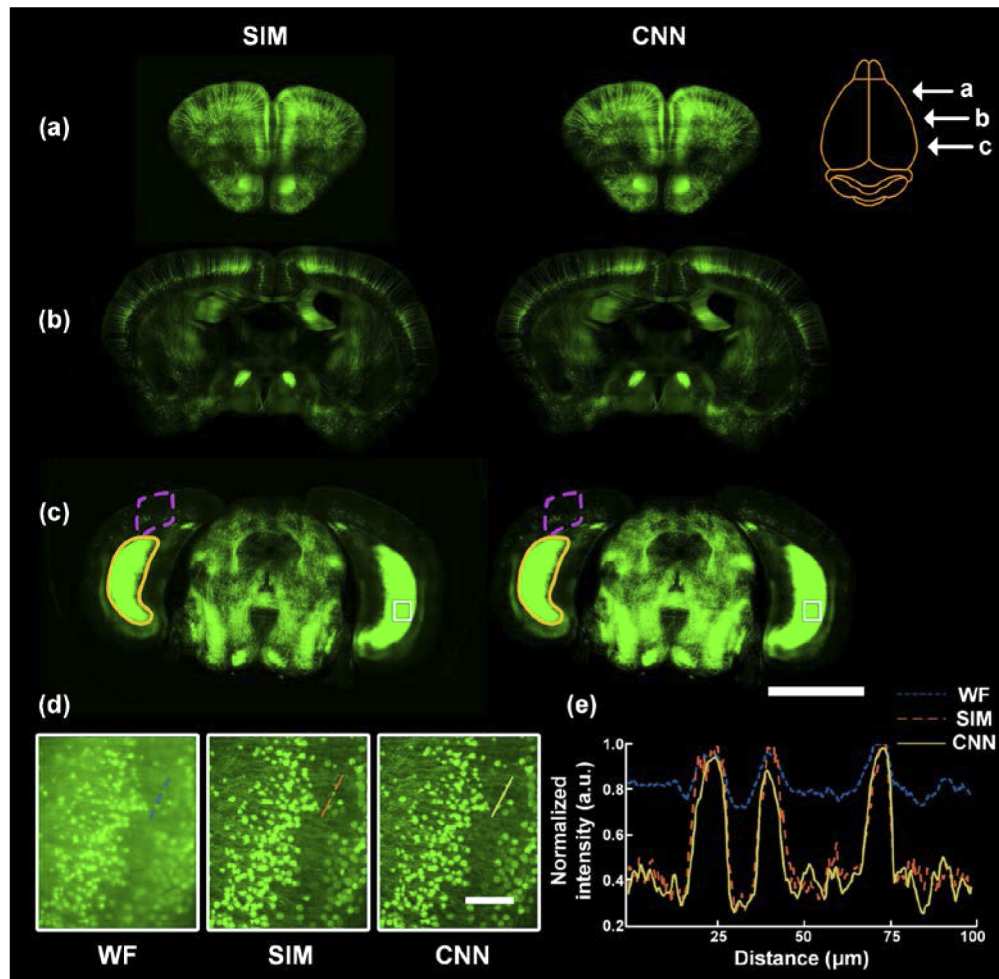


Fig. 4. DL-fMOST imaging of a Thy1-GFP M line mouse brain. (a)–(c) 200- μm -thick MIPs of different coronal sections (2 mm interval) reconstructed by SIM algorithm and CNN prediction. Arrows at the top right corner indicate the locations of the coronal sections. The colored lines in (c) mark two representative brain regions of remarkably different neuron distributions. (d) Enlarged views of the white rectangles in (c). The WF image is shown for comparison. (e) The intensity profiles on color lines of the corresponding images. Scale bar, 2 mm (100 μm for the inset).

solid lines), yet they were sparsely distributed in the primary visual area (highlighted by pink dashed lines). Despite the complex and diverse data characteristics, our deep network performed the optical-sectioning predictions well in different brain regions, compared with the ground truth of the SIM reconstructions. For better illustration, enlarged views of WF, CNN, and SIM reconstructed images of the area marked by the white rectangles are demonstrated (Fig. 4(d)). Normalized pixel intensities along the random selected color lines in the WF, SIM and CNN images were plotted (Fig. 4(e)). Furthermore, we analyzed and compared the SBRs of these images. The SBR value of the WF image was 1.26. And for the SIM and CNN images, the SBR values were 2.14 and 2.27, respectively. These results demonstrated that DL-fMOST had comparable optical sectioning capacity as the WVT system using SIM algorithm.

To evaluate the performance of our network over the whole dataset, we randomly selected a mosaic, took one image every 100 μm along the z direction, and quantified the distance between their CNN outputs and the corresponding ground truths by the RMSE and SSIM. The statistical results are shown in Fig. 5. We tested 100 images in total; all images had RMSE value below 0.025 and SSIM value above 0.85. The average RMSE value was 0.012 and the average SSIM value was 0.92. These results demonstrate that our network performs well across the whole dataset with different image features.

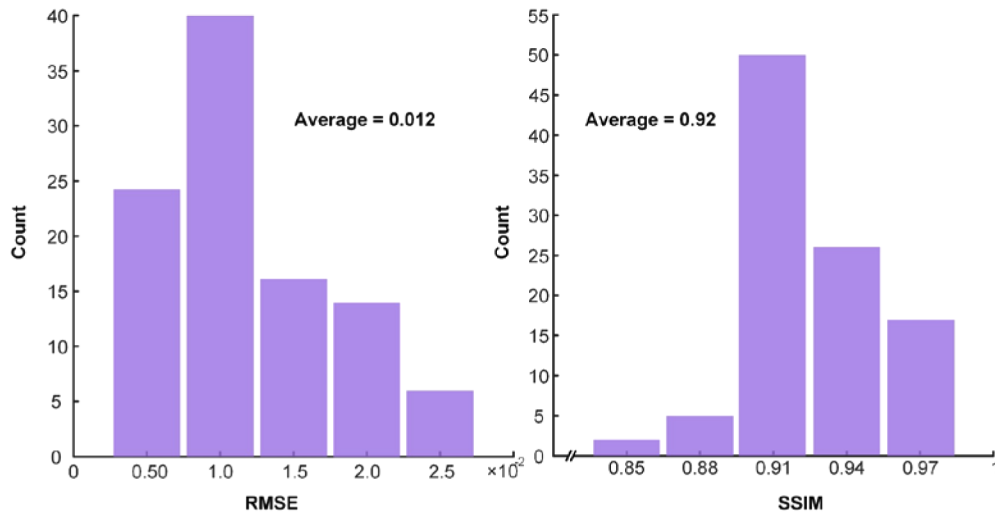


Fig. 5. Quantitative performance of DL-fMOST across the whole-brain dataset. RMSE, root mean squared error; SSIM, structural similarity index.

The acquisition time was quantitatively compared between the original wide-field large-volume tomography (WVT) [20] and DL-fMOST. The data acquisition time for a whole mouse brain consisted of imaging time and sectioning time. We acquired 4 μm z stack with a 2- μm z step and subsequently removed imaged tissue at a sectioning thickness of 4 μm . A typical dataset included approximately 280,000 image stacks. In the WVT system, the imaging time was determined by the switching time of the DMD, movement time of the 3D stage, axial scanning time of the PZT, online running time of the acquisition software, and recording time of the camera. In each field of view (FOV), three phase-shifted raw images were acquired for SIM reconstruction. Average time for each z stack was 780 ms and total imaging time was 60 h. As a contrast, DL-fMOST only acquired one image for each FOV and the system was simplified by the removal of the DMD. The average time for each z stack was optimized to 262 ms, and the total imaging time was reduced to 20 h.

3.2. Continuous image acquisition with high resolution

To verify the integrity of the data in three dimensions (3D), we reconstructed a small volume ($400 \times 400 \times 400 \mu\text{m}$) located in the thalamus in the Thy1-GFP-M mouse brain dataset (Fig. 6(a)) and generated the MIPs along three coordinate directions (Fig. 6(b), Fig. 6(c)). Two profiles on the XZ (Fig. 6(d)) and YZ planes (Fig. 6(e)) were randomly selected and plotted to quantitatively demonstrate the consistency between the SIM images and CNN images. Although the two curves were not exactly identical, the slight difference was mainly reflected in fiber brightness. The CNN prediction did not lose the signal of the neural fibers. To quantify the resolution, we randomly selected 9 neural fibers from the MIPs of the XZ and YZ planes [indicated by red arrows in (Fig. 6(d) and Fig. 6(e))] and measured their full widths at half maximum (FWHM). The average FWHMs calculated from the SIM and CNN images were $3.59 \pm 0.69 \mu\text{m}$ ($n=9$) and $3.69 \pm 0.75 \mu\text{m}$ ($n=9$) for the XZ plane, respectively. And for YZ plane, the results were $3.70 \pm 0.71 \mu\text{m}$ ($n=9$) and $3.84 \pm 0.81 \mu\text{m}$ ($n=9$), respectively. These quantitative analyses

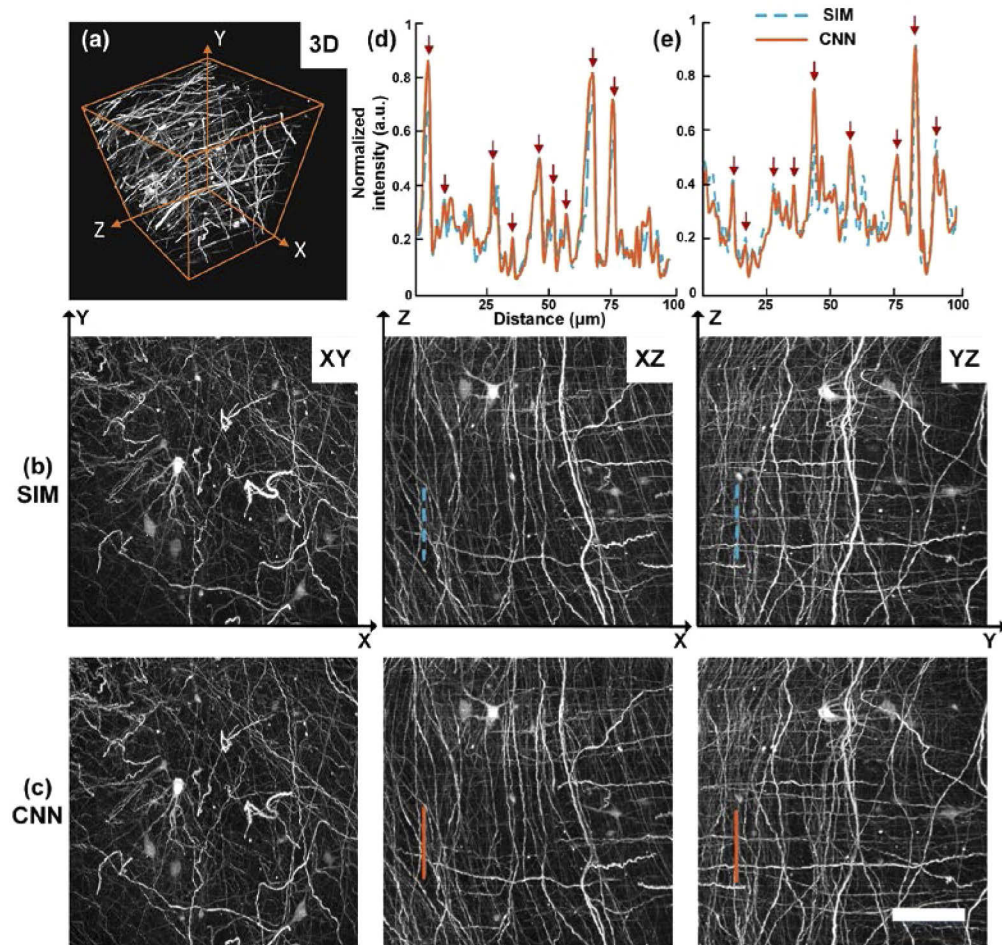


Fig. 6. 3D imaging capability of DL-fMOST. (a) 3D reconstruction of a data block extracted from the test set 1. (b,c) MIPs of the SIM volume and CNN volume in three coordinate directions. The intensity profiles are plotted according to the colored dashed lines on the xz (d) and yz (e) planes, where the red arrows indicate the local peaks of the neural fibers. Scale bar: 100 μm .

verified that DL-fMOST had comparable image quality as the WVT system based on SIM, which provided a guarantee for subsequent downstream analyses.

Furthermore, we compared the data integrities of CNN output and SIM ground truth for 3D reconstruction (Fig. 7). In comparison with the SIM images (Fig. 7(a)), the CNN output images (Fig. 7(b)) had similar contrast and resolution. We can clearly distinguish individual axons from the dense fiber bundles, and even for weak fiber signals, the network output still matched well with the SIM ground truth (insets in Fig. 7). We merged the SIM images with the CNN output images to further investigate the performance of DL-fMOST (Fig. 7(c)). We found that the difference was mainly reflected in the brightness of the edge area (indicated by white arrows in Fig. 7). In SIM images, the signal intensities at the corner of the FOV were relatively lower, which was caused by gaussian wide-field illumination. Our DL-fMOST leveraged batch

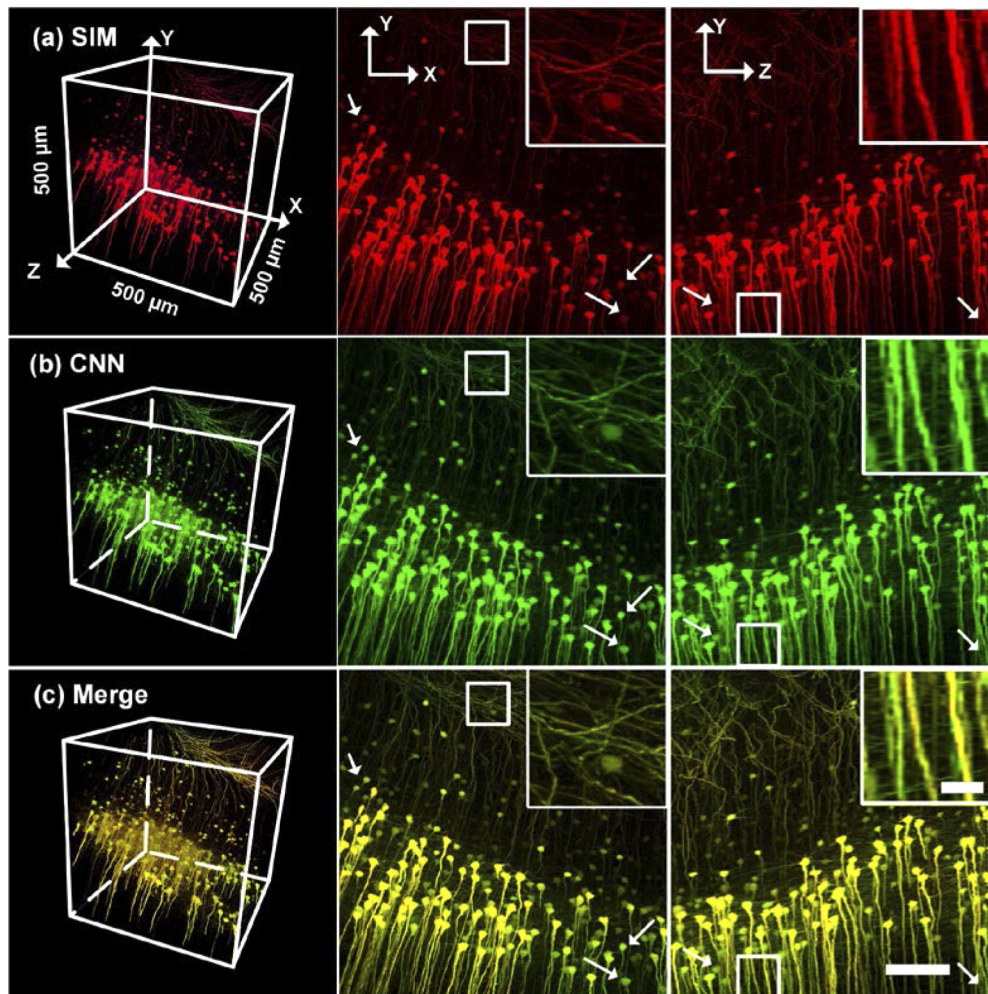


Fig. 7. 3D reconstruction of a GFP-labeled image stack and MIPs in the XY direction and YZ direction using (a) SIM algorithm and (b) CNN. (c) Merged image of (a) and (b), the mapped areas are shown in yellow. The white arrows indicate that the signal in the corner of the FOV will be enhanced by CNN. The insets show the magnifications of the areas marked with white solid lines. The image block is selected from the same whole-brain dataset as in Fig. 4. Scale bar, 100 μm (20 μm in the inset).

normalization technique [39] in the training phase, which potentially normalized the brightness of captured images, resulting in a more uniform light field in the reconstructed images. Thus, the weak signals at the corner of the FOV were easier to distinguish.

3.3. SNR enhancement and artifact removal

To evaluate noise reduction and SNR improvement benefiting from deep learning in our method, we compared the noise levels of a pair of SIM and CNN reconstructed images from test set 1, as shown in Fig. 8. There was strong background noise on the SIM-reconstructed image (Fig. 8(a)). The background was cleaner in the CNN-reconstructed image (Fig. 8(b)). The SNRs of SIM- and CNN-reconstructed images were 9.30 and 62.03, respectively. This benefits from the noise rejection of deep learning. Because the noise is random and unpredictable, the network learns to output the average of all plausible explanations for minimizing the overall loss between the network output and the true target. The main noise in the raw images, such as white Gaussian noise and Poisson noise, are zero-mean, so the network learns to output the clean image. Our finding was consistent with Lehtinen et al. [45]. In SIM imaging, imperfect illumination modulation may cause the appearance of stripe artifacts. Additionally, the modulation is sample dependent and spatially varying [46]. Therefore, we can observe the residual stripe artifacts in some areas of the sample. By leveraging deep learning to directly predict optical-sectioned images from WF input, such artifacts can be effectively reduced.

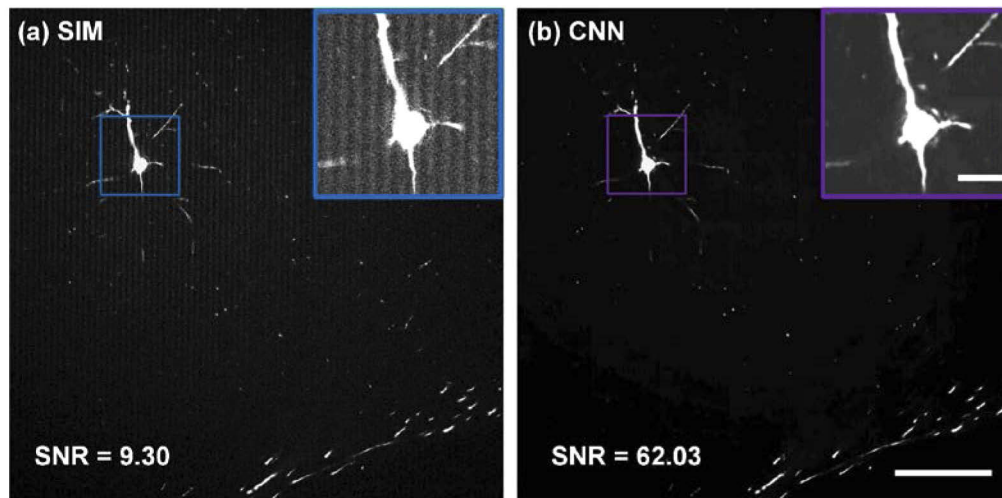


Fig. 8. SNR improvement and artifacts elimination in DL-fMOST. (a) SIM-reconstructed image and (b) CNN-reconstructed image. Both images had been given the same contrast stretch to improve the display. Scale bar, 100 μm (20 μm in the inset)

3.4. Applicability of DL-fMOST

To test the applicability of DL-fMOST to different types of neuronal morphology and distributions, we reconstructed four whole-brain datasets (test set 2-5) with various fluorescence-labeled neuron types using the SIM algorithm and CNN prediction. We directly used the Thy1-GFP-M trained model for inference without retraining or transfer learning. Figure 9 shows the typical coronal section near the injection site based on CNN reconstruction of each dataset. There are significant characteristic differences in the neuron morphology and distribution patterns among the four samples. The predictions of the CNN network still enables accurate reconstruction of both the cell body morphology and axon projections. The enlarged views of the white rectangles in Fig. 9

further show the precise match between CNN output images and corresponding SIM ground-truth images. We further quantified the reconstruction performance of DL-fMOST compared with the SIM images. The average SSIM and RMSE were 0.9044 and 0.0166, respectively. Our results showed the robustness of DL-fMOST in the various sample types and changes of acquisition parameters.

3.5. Soma localization

To validate the capability of our method to support cell identification in 3D, we performed and compared stereological cell locating and counting on the test set 2. We randomly selected 10 data blocks of $200 \times 200 \times 200 \mu\text{m}$ near the injection site and used NeuroGPS software [47] to automatically count the soma. Figure 10(a) and Fig. 10(b) show the representative volume reconstructed by SIM algorithm and CNN, respectively. 51 soma centers were automatically identified in SIM data (indicated by the red dots), and 53 in CNN data. There were three missed soma centers in the SIM block, and they were accurately identified in the CNN block (indicated by the blue arrows). The green arrow points to the soma accurately identified in the SIM block that were missed in the CNN block. Two background signals were mistakenly identified in the SIM block and one in CNN block (indicated by the orange arrows). In both data blocks, one cell body near the bounding box was not detected (indicated by the purple arrows). In order to validate the accuracy of the results, we compared the counting results with the manual identification and calculated the precision rate and recall rate, as shown in Fig. 10(c) and Fig. 10(d). Average precision and recall rate were $96.3\% \pm 1.6\%$ ($n = 10$) and $94.6\% \pm 2.4\%$ ($n = 10$) for CNN data, and $96.2\% \pm 1.1\%$ ($n = 10$) and $92.8\% \pm 2.7\%$ ($n = 10$) for SIM data. The counting accuracy of CNN data was slightly higher than that of SIM data. This benefits from that the higher SNR of the CNN output images, which is more conducive to the separation of foreground and background signals.

3.6. Single neuron reconstruction

To further demonstrate the high-resolution, high-contrast, and continuous 3D imaging capacities of DL-fMOST, we performed single-neuron morphology reconstruction on the CNN output image stack and the corresponding SIM ground-truth image stack from the test set 3. A data volume ($500 \times 600 \times 1000 \mu\text{m}$) located in the neocortex was extracted for validation as shown in Fig. 11(a). Figure 11(b) and Fig. 11(c) show the 200- μm -thick MIPs of the SIM data block and CNN data block, which indicate the target neurons. We employed a semi-automated pipeline for fast and accurate morphology reconstruction. Specifically, a coarse tracing result was automatically done by the NeuroGPS-Tree algorithm [48]. Next, the human annotators inspected the neuronal tree and corrected the wrong connections as well as completed the missing branches. The fiber tracing experiment was double-blinded to ensure unbiased analysis by two skilled annotators. The semi-automated tracing results are shown in Fig. 11(d). The neuron morphology reconstructed from the DL-fMOST generated image stack and SIM ground-truth image stack were almost the same. Using SIM reconstruction results as the gold standard, the recall and precision rates for CNN reconstructed block were 99.27% and 99.12%, respectively. In addition, we derived the total lengths and branch numbers from the tracing results of the SIM and CNN reconstruction. The total lengths of the neural fibers were 16.03 mm and 16.02 mm. Both the SIM and CNN reconstruction had the same branch numbers, which were 81. These results demonstrated the DL-fMOST had comparable performance in single neuron reconstruction as the WVT technology based on SIM imaging.

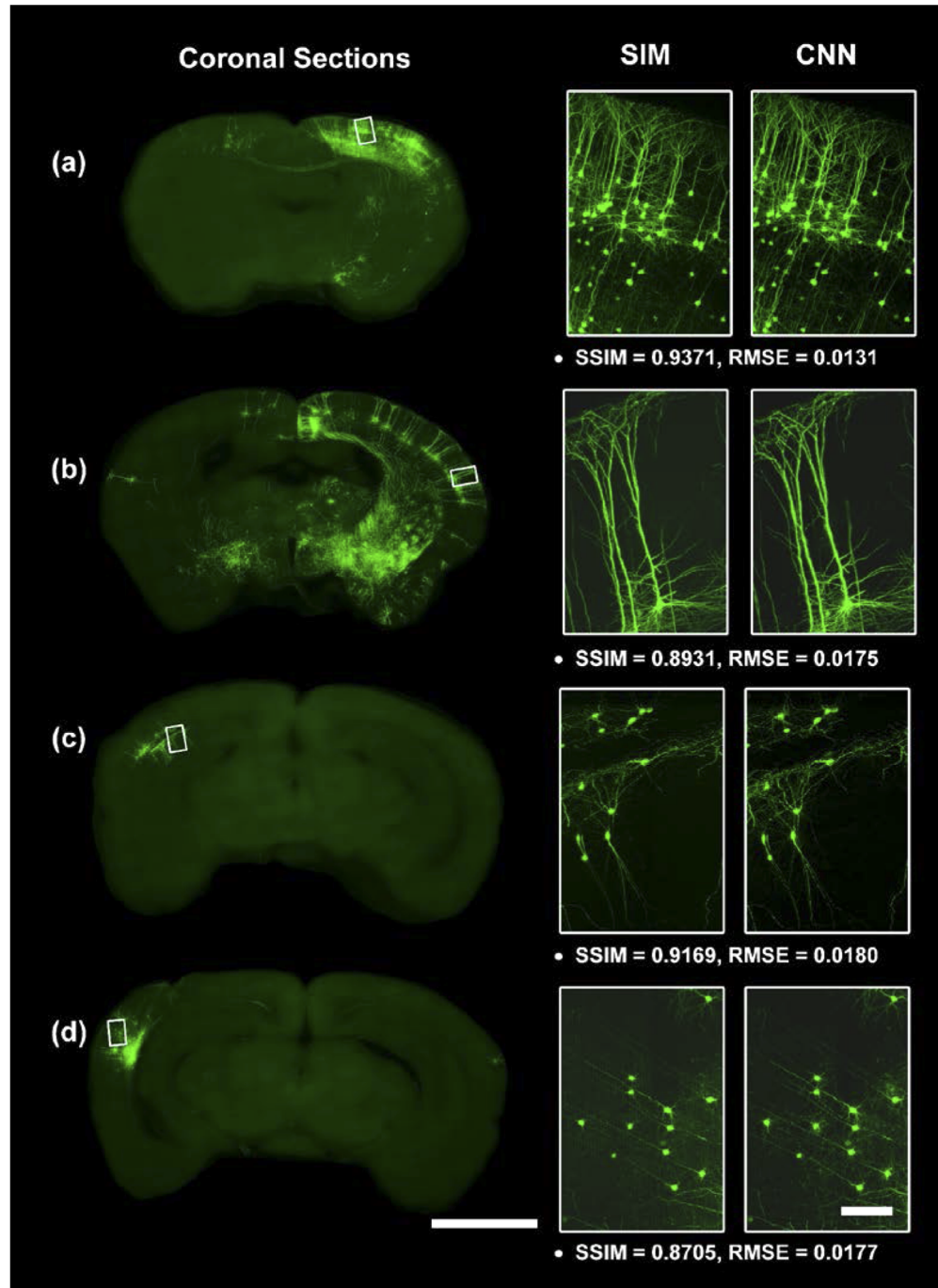


Fig. 9. Imaging mouse brains of different labeling targets. (a)–(d) Typical coronal MIPs of four mouse brains (test set 2-5) with different projection patterns reconstructed by the same Thy1-GFP-M trained CNN. The MIP thicknesses are 200 μm . Scale bar, 2 mm (100 μm for the inset).

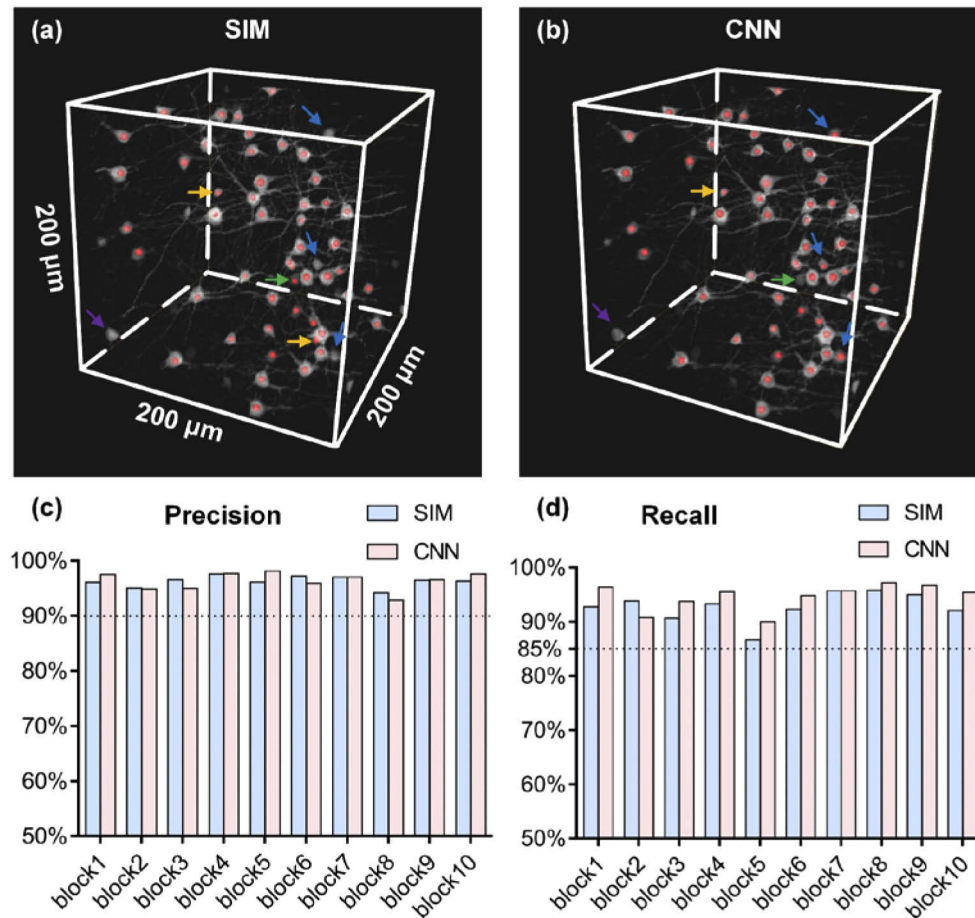


Fig. 10. Quantitative cell counting comparison. Automatically locating soma centers in the (a) SIM data and (b) CNN data using NeuroGPS algorithm (red dots). Blue arrows indicate cell bodies that were not correctly identified in the SIM block, but accurately identified in the CNN block. The green arrow marks the opposite of the blue arrows. Erroneous identifications are indicated by orange arrows. All missed somas in both data blocks are marked by purple arrows. (c) Precision and (d) recall rates for both SIM and CNN image stacks.

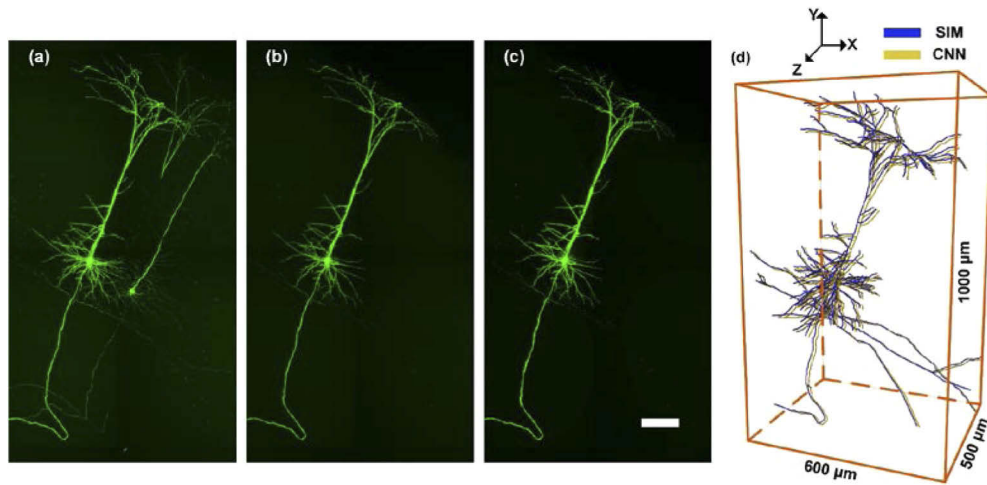


Fig. 11. Neuron morphological reconstruction comparison. (a) 500 μm MIPs of the image stack along the Z direction. (b)-(c) 200 μm MIPs derived from the SIM and CNN reconstruction. (d) Semi-automated tracing results based on (b, c). Scale bar, 100 μm .

4. Discussion and conclusion

Here, we present DL-fMOST, a novel deep-learning-based whole-brain imaging method, high-throughput and high-resolution. An important feature of our method is the use of a CNN to implement real-time optical sectioning, which greatly reduces the complexity of the imaging system and improves stability and acquisition speed. We only need to build the simplest wide-field fluorescence microscope, coupled with the sectioning module, to achieve whole-brain mapping at a voxel size of $0.32 \times 0.32 \times 2 \mu\text{m}$ in 1.5 d. Another benefit brought by deep learning is the capacity to suppress noise and remove artifacts, which means we can directly use the raw data without cumbersome post-processing.

Recently, some laboratories have developed some solutions to significantly shorten the whole-brain imaging time to several hours [21,22]. These methods use low magnification objectives with compromised 3D resolutions to quickly acquire cell distribution throughout the brains. However, their imaging qualities were not good enough to distinguish individual neural fibers and reconstruct neuronal morphology. Our DL-fMOST method generates high-quality whole brain dataset at a submicron voxel size and enables counting of cell distribution and tracing neuronal morphology. Currently, the imaging throughput of this system is limited by the frame rate of the camera. With an objective of larger FOV and a detector of larger chip size, the acquisition speed will be further improved in the future. In addition, deep learning can not only realize optical sectioning, but also be expected to combine with subsequent data analysis functions such as cell counting and fiber tracing to realize online data acquisition and analysis in the future. DL-fMOST generalizes well to various types of the neurons without retraining. However, we also need to note that current deep learning methods generally have poor performance of test data outside of distribution [24]. Therefore, DL-fMOST needs to be used after retraining when it is used for structures with very different anatomical characteristics, such as cytoarchitecture and vascular network. In the future, the neural network may be further upgraded and optimized to construct a universal network for various types of structural features.

In summary, DL-fMOST has the potential to facilitate exploration of neuron populations and neural circuits at single-axon resolution, providing new tools for understanding cell types and connectivity across the brain.

Funding

National Key Research and Development Program of China (2017YFA0700402); National Natural Science Foundation of China (81671374, 91749209, 92032000).

Acknowledgments

We thank the members of the MOST group from the Britton Chance Centre for Biomedical Photonics for their assistance with experiments and comments on the manuscript. We appreciate Prof. Xin Yang for valuable suggestions. We thank the Optical Bioimaging Core Facility of WNLO-HUST for the support of data acquisition.

Disclosures

The authors declare no conflicts of interest.

References

1. J. W. Lichtman and W. Denk, "The big and the small: Challenges of imaging the brain's circuits," *Science* **334**(6056), 618–623 (2011).
2. B. Zingg, H. Hintiryan, L. Gou, M. Y. Song, M. Bay, M. S. Bienkowski, N. N. Foster, S. Yamashita, I. Bowman, A. W. Toga, and H.-W. Dong, "Neural networks of the mouse neocortex," *Cell* **156**(5), 1096–1111 (2014).
3. B. J. Hunnicutt, B. R. Long, D. Kusefoglou, K. J. Gertz, H. Zhong, and T. Mao, "A comprehensive thalamocortical projection map at the mesoscopic level," *Nat. Neurosci.* **17**(9), 1276–1285 (2014).
4. S. W. Oh, J. A. Harris, L. Ng, B. Winslow, N. Cain, S. Mihalas, Q. Wang, C. Lau, L. Kuan, A. M. Henry, M. T. Mortrud, B. Ouellette, T. N. Nguyen, S. A. Sorensen, C. R. Slaughterbeck, W. Wakeman, Y. Li, D. Feng, A. Ho, E. Nicholas, K. E. Hirokawa, P. Bohn, K. M. Joines, H. Peng, M. J. Hawrylycz, J. W. Phillips, J. G. Hohmann, P. Wohnoutka, C. R. Gerfen, C. Koch, A. Bernard, C. Dang, A. R. Jones, and H. Zeng, "A mesoscale connectome of the mouse brain," *Nature* **508**(7495), 207–214 (2014).
5. G. M. G. Shepherd, M. Raastad, and P. Andersen, "General and variable features of varicosity spacing along unmyelinated axons in the hippocampus and cerebellum," *Proc. Natl. Acad. Sci. U. S. A.* **99**(9), 6340–6345 (2002).
6. Y. Sun, A. Q. Nguyen, J. P. Nguyen, L. Le, D. Saur, J. Choi, E. M. Callaway, and X. Xu, "Cell-type-specific circuit connectivity of hippocampal CA1 revealed through Cre-dependent rabies tracing," *Cell Rep.* **7**(1), 269–280 (2014).
7. D. Zhu, K. V. Larin, Q. Luo, and V. V. Tuchin, "Recent progress in tissue optical clearing," *Laser Photonics Rev.* **7**(5), 732–757 (2013).
8. H.-U. Dodt, U. Leischner, A. Schierloh, N. Jähring, C. P. Mauch, K. Deininger, J. M. Deussing, M. Eder, W. Ziegglänsberger, and K. Becker, "Ultramicroscopy: three-dimensional visualization of neuronal networks in the whole mouse brain," *Nat. Methods* **4**(4), 331–336 (2007).
9. K. Chung and K. Deisseroth, "CLARITY for mapping the nervous system," *Nat. Methods* **10**(6), 508–513 (2013).
10. E. A. Susaki, K. Tainaka, D. Perrin, F. Kishino, T. Tawara, T. M. Watanabe, C. Yokoyama, H. Onoe, M. Eguchi, S. Yamaguchi, T. Abe, H. Kiyonari, Y. Shimizu, A. Miyawaki, H. Yokota, and H. R. Ueda, "Whole-brain imaging with single-cell resolution using chemical cocktails and computational analysis," *Cell* **157**(3), 726–739 (2014).
11. N. Renier, E. L. Adams, C. Kirst, Z. Wu, R. Azevedo, J. Kohl, A. E. Autry, L. Kadiri, K. Umadevi Venkataraju, Y. Zhou, V. X. Wang, C. Y. Tang, O. Olsen, C. Dulac, P. Osten, and M. Tessier-Lavigne, "Mapping of brain activity by automated volume analysis of immediate early genes," *Cell* **165**(7), 1789–1802 (2016).
12. J. Yuan, H. Gong, A. Li, X. Li, S. Chen, S. Zeng, and Q. Luo, "Visible rodent brain-wide networks at single-neuron resolution," *Front. Neuroanat.* **9**, 70 (2015).
13. A. Li, H. Gong, B. Zhang, Q. Wang, C. Yan, J. Wu, Q. Liu, S. Zeng, and Q. Luo, "Micro-optical sectioning tomography to obtain a high-resolution atlas of the mouse brain," *Science* **330**(6009), 1404–1408 (2010).
14. H. Gong, S. Zeng, C. Yan, X. Lv, Z. Yang, T. Xu, Z. Feng, W. Ding, X. Qi, A. Li, J. Wu, and Q. Luo, "Continuously tracing brain-wide long-distance axonal projections in mice at a one-micron voxel resolution," *NeuroImage* **74**, 87–98 (2013).
15. T. Ragan, L. R. Kadiri, K. U. Venkataraju, K. Bahlmann, J. Sutin, J. Taranda, I. Arganda-Carreras, Y. Kim, H. S. Seung, and P. Osten, "Serial two-photon tomography for automated ex vivo mouse brain imaging," *Nat. Methods* **9**(3), 255–258 (2012).
16. T. Zheng, Z. Yang, A. Li, X. Lv, Z. Zhou, X. Wang, X. Qi, S. Li, Q. Luo, H. Gong, and S. Zeng, "Visualization of brain circuits using two-photon fluorescence micro-optical sectioning tomography," *Opt. Express* **21**(8), 9839–9850 (2013).
17. M. N. Economo, N. G. Clack, L. D. Lavis, C. R. Gerfen, K. Svoboda, E. W. Myers, and J. Chandrashekar, "A platform for brain-wide imaging and reconstruction of individual neurons," *eLife* **5**, e10566 (2016).
18. J. Winnubst, E. Bas, T. A. Ferreira, Z. Wu, M. N. Economo, P. Edson, B. J. Arthur, C. Bruns, K. Rokicki, D. Schauder, D. J. Olbris, S. D. Murphy, D. G. Ackerman, C. Arshadi, P. Baldwin, R. Blake, A. Elsayed, M. Hasan, D. Ramirez,

- B. Dos Santos, M. Weldon, A. Zafar, J. T. Dudman, C. R. Gerfen, A. W. Hantman, W. Korff, S. M. Sternson, N. Spruston, K. Svoboda, and J. Chandrashekar, "Reconstruction of 1,000 projection neurons reveals new cell types and organization of long-range connectivity in the mouse brain," *Cell* **179**(1), 268–281.e13 (2019).
19. M. A. A. Neil, R. Juskaitis, and T. Wilson, "Method of obtaining optical sectioning by using structured light in a conventional microscope," *Opt. Lett.* **22**(24), 1905–1907 (1997).
 20. H. Gong, D. Xu, J. Yuan, X. Li, C. Guo, J. Peng, Y. Li, L. A. Schwarz, A. Li, B. Hu, B. Xiong, Q. Sun, Y. Zhang, J. Liu, Q. Zhong, T. Xu, S. Zeng, and Q. Luo, "High-throughput dual-colour precision imaging for brain-wide connectome with cytoarchitectonic landmarks at the cellular level," *Nat. Commun.* **7**(1), 12142 (2016).
 21. K. Seiriki, A. Kasai, T. Hashimoto, W. Schulze, M. Niu, S. Yamaguchi, T. Nakazawa, K.-i. Inoue, S. Uezono, M. Takada, Y. Naka, H. Igarashi, M. Tanuma, J. A. Waschek, Y. Ago, K. F. Tanaka, A. Hayata-Takano, K. Nagayasu, N. Shintani, R. Hashimoto, Y. Kunii, M. Hino, J. Matsumoto, H. Yabe, T. Nagai, K. Fujita, T. Matsuda, K. Takuma, A. Baba, and H. Hashimoto, "High-speed and scalable whole-brain imaging in rodents and primates," *Neuron* **94**(6), 1085–1100.e6 (2017).
 22. X. Yang, Q. Zhang, F. Huang, K. Bai, Y. Guo, Y. Zhang, N. Li, Y. Cui, P. Sun, S. Zeng, and X. Lv, "High-throughput light sheet tomography platform for automated fast imaging of whole mouse brain," *J. Biophotonics* **11**(9), e201800047 (2018).
 23. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* **521**(7553), 436–444 (2015).
 24. C. Belthangady and L. A. Royer, "Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction," *Nat. Methods* **16**(12), 1215–1225 (2019).
 25. K. de Haan, Y. Rivenson, Y. Wu, and A. Ozcan, "Deep-learning-based image reconstruction and enhancement in optical microscopy," *Proc. IEEE* **108**(1), 30–50 (2020).
 26. X. Zhang, Y. Chen, K. Ning, C. Zhou, Y. Han, H. Gong, and J. Yuan, "Deep learning optical-sectioning method," *Opt. Express* **26**(23), 30762–30772 (2018).
 27. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
 28. Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," *Optica* **4**(11), 1437–1443 (2017).
 29. Y. Wu, Y. Luo, G. Chaudhari, Y. Rivenson, A. Calis, K. de Haan, and A. Ozcan, "Bright-field holography: cross-modality deep learning enables snapshot 3D imaging with bright-field contrast using a single hologram," *Light: Sci. Appl.* **8**(1), 25 (2019).
 30. G. Feng, R. H. Mellor, M. Bernstein, C. Keller-Peck, Q. T. Nguyen, M. Wallace, J. M. Nerbonne, J. W. Lichtman, and J. R. Sanes, "Imaging Neuronal Subsets in Transgenic Mice Expressing Multiple Spectral Variants of GFP," *Neuron* **28**(1), 41–51 (2000).
 31. W. Ouyang, A. Aristov, M. Lelek, X. Hao, and C. Zimmer, "Deep learning massively accelerates super-resolution localization microscopy," *Nat. Biotechnol.* **36**(5), 460–468 (2018).
 32. H. Wang, Y. Rivenson, Y. Jin, Z. Wei, R. Gao, H. Günaydin, L. A. Bentolila, C. Kural, and A. Ozcan, "Deep learning enables cross-modality super-resolution in fluorescence microscopy," *Nat. Methods* **16**(1), 103–110 (2019).
 33. C. Ounkomol, S. Seshamani, M. M. Maleckar, F. Collman, and G. R. Johnson, "Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy," *Nat. Methods* **15**(11), 917–920 (2018).
 34. Y. Rivenson, T. Liu, Z. Wei, Y. Zhang, K. de Haan, and A. Ozcan, "PhaseStain: the digital staining of label-free quantitative phase microscopy images using deep learning," *Light: Sci. Appl.* **8**(1), 23 (2019).
 35. B. Manifold, E. Thomas, A. T. Francis, A. H. Hill, and D. Fu, "Denoising of stimulated Raman scattering microscopy images via deep learning," *Biomed. Opt. Express* **10**(8), 3860–3874 (2019).
 36. M. Weigert, U. Schmidt, T. Boothe, A. Müller, A. Dibrov, A. Jain, B. Wilhelm, D. Schmidt, C. Broadus, S. Culley, M. Rocha-Martins, F. Segovia-Miranda, C. Norden, R. Henriques, M. Zerial, M. Solimena, J. Rink, P. Tomancak, L. Royer, F. Jug, and E. W. Myers, "Content-aware image restoration: pushing the limits of fluorescence microscopy," *Nat. Methods* **15**(12), 1090–1097 (2018).
 37. A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proceedings of the 30th International Conference on Machine Learning* (2013), pp. 3–8.
 38. X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics* (2011), pp. 315–323.
 39. S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," arXiv preprint arXiv:1502.03167 (2015).
 40. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.* **15**(1), 1929–1958 (2014).
 41. H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for neural networks for image processing," arXiv preprint arXiv:1511.08861 (2015).
 42. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," arXiv preprint arXiv:1412.6980 (2014).
 43. A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in PyTorch," in *31st Conference on Neural Information Processing Systems (NIPS)* (2017).
 44. Z. Wang, A. C. Bovik, and L. Lu, "Why is image quality assessment so difficult?" in: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* (IEEE, 2002), vol. 4, pp. 3313–3316.

45. J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2Noise: learning image restoration without clean data," arXiv preprint arXiv:1803.04189 (2018).
46. L. H. Schaefer, D. Schuster, and J. Schaffer, "Structured illumination microscopy: artefact analysis and reduction utilizing a parameter optimization approach," *J. Microsc.* **216**(2), 165–174 (2004).
47. T. Quan, T. Zheng, Z. Yang, W. Ding, S. Li, J. Li, H. Zhou, Q. Luo, H. Gong, and S. Zeng, "NeuroGPS: automated localization of neurons for brain circuits using L1 minimization model," *Sci. Rep.* **3**(1), 1414 (2013).
48. T. Quan, H. Zhou, J. Li, S. Li, A. Li, Y. Li, X. Lv, Q. Luo, H. Gong, and S. Zeng, "NeuroGPS-Tree: automatic reconstruction of large-scale neuronal populations with dense neurites," *Nat. Methods* **13**(1), 51–54 (2016).